A conversation with Professor Anthony Zador, September 12, 2019

Participants

- Professor Anthony Zador - Alle Davis and Maxine Harrison Professor of Neurosciences, Cold Spring Harbor Laboratory
- Joseph Carlsmith - Research Analyst, Open Philanthropy

Note: These notes were compiled by Open Philanthropy and give an overview of the major points made by Prof. Zador.

Summary

Open Philanthropy spoke with Prof. Anthony Zador of Cold Spring Harbor Laboratory as part of its investigation of what we can learn from the brain about the computational power ("compute") sufficient to match human-level task performance. The conversation focused on the computational role of different processes in the brain.

Computation and biological detail

Understanding biological details in the brain is crucially important to understanding neural mechanisms, disease, drug design, and much else. When it comes to replicating human intelligence, however, the point of understanding such mechanisms is to be able to step back and identify what's important.

As an analogy, suppose you wanted to build a car, using an existing car as a template. There are certain key features -- internal combustion engine, wheels, etc -- that you'll probably want to carry over. But many details, like the thread count on the bolts, don't need to be replicated. However, if you want to *fix* an existing car, you need exactly the right bolt.

Neuron modeling

Prof. Zador believes that integrate-and-fire neuron models, or something like them, are adequate to capture the contribution of a neuron to the brain's information-processing. He does not think that Hodgkin-Huxley-type models are required, or that we need to include the details of synaptic conductances in our models.

However, he believes that the temporal dynamics of spiking are important. That is, it matters that there are discrete spikes, occurring at particular moments in time, which are the conduit of information between neurons. There have been some attempts to build artificial neural networks with spiking neurons. In current work, spikes generally seem more like a bug than a feature, but Prof. Zador believes that spikes are actually a natural way to implement a device that operates in real time, which most neural networks don't.

That said, he does not think that the nuances of *how* these spikes are generated matter very much. The integrate and fire model is one mathematically tractable model, but there are others which, if more mathematically tractable, would be fine as well.

Interchangeable non-linearities

In the early days of artificial neural networks, people thought that sigmoid activation functions were required, and that piecewise linear models could not work because they are not differentiable. But it turns out that computers can handle the activation function having one non-differentiable point, so the two are largely interchangeable, and it's fine to go with the more convenient option. The main constraint is that the function needs to be monotonically increasing.

This is an example of a case in which the precise function generating a neuron-like unit's output does not matter.

Dynamical synapses

Real synapses in the brain are dynamic: their behavior depends on previous activity at that synapse. This has never really percolated into artificial neural network theory, but Prof. Zador believes that this is a historical fluke, and that dynamic synapses could be a rich area to explore. He wrote a paper a few years ago on the topic.

There is a big difference, computationally, between processes that happen at every synapse, and processes that only happen at the soma, because there are orders of magnitude fewer somas than synapses. Synapses are the tunable parameters in the brain.

Mechanisms that span timescales

You need a handful of mechanisms that span timescales, and that go beyond the timescale of a spike or a synaptic current or a membrane time constant, in order to capture things like persistent memory. Possibilities here include NMDA currents, short-term synaptic plasticity (Prof. Zador's favorite candidate), reverberating circuits, and spike generating mechanisms with some latent variable (some calcium currents in the thalamus are an example of this). There are other possibilities as well.

Dendritic computation

Much of Prof. Zador's PhD work was devoted to the hypothesis that dendritic computation is the key difference between artificial neural networks and real brains. However, at the end of the day, he was led to the conclusion that dendritic computation does not make a qualitative difference to the computational capacity of a neuron. There is some computational boost, but the same effect could be achieved by replacing each biological neuron with a handful of artificial neurons.

Learning

A lot of the artificial neural network community is focused on learning, but Prof. Zador believes that a lot of what matters for intelligence is wired into the brain by evolution, rather than learned (a view he thinks fairly obvious to those who work on non-human animals).

What's impressive about mice is not what they learn, but what they can do prior to learning. Indeed, it can be frustrating to work with mice, because it is difficult to teach them to perform tasks, and you have to find a method of training that fits with their natural methods of learning. Human learning is clearly impressive, but Prof. Zador expects that if we could create mouse-level intelligence, the central foundations would be in place, and moving to human-level intelligence would be fairly easy. After all, human learning evolved relatively recently on evolutionary timescales.

Compute for synaptic plasticity

We know the general outlines of the rules governing synaptic plasticity. The synapse gets stronger and weaker as a function of pre and post synaptic activity, and external modulation. There is a lot of room for discovery there, and it may be difficult to get just right, but conceptually, it's fairly simple. Prof. Zador expects it to be possible to capture synaptic plasticity with a small number of FLOPs per spike through synapse.

There are various proposals for implementing backpropagation in the brain, but Prof. Zador is skeptical of these.

Alternative signaling mechanisms

- Glia are very important to understanding disease, but Prof. Zador thinks that they are unlikely to be important to computing in the brain.
- Prof. Zador believes that neuromodulation is the dominant form of global signaling in the brain. However, while global signals may be very important to a model's function, they won't add much computational burden (the same goes for processes that proceed on longer timescales). It takes fewer bits to specify a global signal, almost by definition.
- Prof. Zador believes that ephaptic communication is unlikely to be important to the brain's information-processing. Even if it was important, though, it would be a form of global signaling, and so comparatively inexpensive to model.

Unknown unknowns

Prof. Zador is skeptical that there are major unknown unknowns in the parts list in the brain, given how much effort has gone into studying nervous systems. Biology is complicated, and there is still more to understand, but Prof. Zador does not think that what we are missing is a breakthrough in biology.

Rather, what's missing is an understanding of the brain's organizing principles. When it comes to understanding neural circuits, there are huge unknown unknowns. Indeed, one possibility is that what's missing is an adequate understanding of the wiring diagram, but that there is no easy conceptual breakthrough that will reveal the important principles -- rather, 500 million years of evolution just created a very clever wiring diagram.

Back of the envelope calculation

Here is a rough, back-of-the-envelope calculation of the compute power sufficient to replicate the human brain's computation. The human cortex has about 1e10 neurons, each with about 1e3 connections, spiking about once per second. This would lead to an overall estimate of 1e13 floating point operations per second, though the true number of spikes and/or synapses might be somewhat different.

Relevance of progress in image recognition

Prof. Zador is impressed by recent progress in computer vision, but he does not think that it provides much evidence about the compute resources necessary to do what the brain does.
This is for the same reason that Deep Blue beating Garry Kasparov did not provide such evidence. Both systems are important and interesting solutions to very specific problems, but they aren't very relevant to the resource requirements for general intelligence.

The vulnerability of ANN vision models to adversarial attack also fits with the view that they aren't performing the task in the same way humans are. That said, there are some parallels between the feature detection performed in the early layers of the visual cortex, and what ANN vision systems are doing (though one may be able to say something analogous about the Deep Blue case).

Prof. Zador shares the commonly held view that embodiment is key to intelligence. We need to build agents that navigate the real world (or possibly, a simulated world) in real time.

Sources of views and disagreements

Prof. Zador's views about the computational role of different neural mechanisms are shaped centrally by gut feeling and scientific aesthetic. Neuroscientists have debated this issue for decades, and ultimately the proof is in the pudding.

Prof. Zador expects that a lot of neuroscientists would say that "we just don't know" what amount of compute would be required to match human-level task performance. There is also a wide diversity of views in the field, and many people's views are centrally shaped by their research background. For example, people with backgrounds in biology are generally more excited about incorporating biological detail; people who study humans tend to focus on learning; and people who study small animals will like C. elegans or fruit flies focus less on learning and more on innate behaviors.